

# Psychiatric Illnesses as Disorders of Network Dynamics

Daniel Durstewitz, Quentin J.M. Huys, and Georgia Koppe

## ABSTRACT

This review provides a dynamical systems perspective on mental illness. After a brief introduction to the theory of dynamical systems, we focus on the common assumption in theoretical and computational neuroscience that phenomena at subcellular, cellular, network, cognitive, and even societal levels could be described and explained in terms of dynamical systems theory. As such, dynamical systems theory may also provide a framework for understanding mental illnesses. The review examines a number of core dynamical systems phenomena and relates each of these to aspects of mental illnesses. This provides an outline of how a broad set of phenomena in serious and common mental illnesses and neurological conditions can be understood in dynamical systems terms. It suggests that the dynamical systems level may provide a central, hublike level of convergence that unifies and links multiple biophysical and behavioral phenomena in the sense that diverse biophysical changes can give rise to the same dynamical phenomena and, vice versa, similar changes in dynamics may yield different behavioral symptoms depending on the brain area where these changes manifest. We also briefly outline current methodological approaches for inferring dynamical systems from data such as electroencephalography, functional magnetic resonance imaging, or self-reports, and we discuss the implications of a dynamical view for the diagnosis, prognosis, and treatment of psychiatric conditions. We argue that a consideration of dynamics could play a potentially transformative role in the choice and target of interventions.

**Keywords:** Attractor, Chaos, Dynamical systems, Machine learning, Recurrent neural networks, Schizophrenia

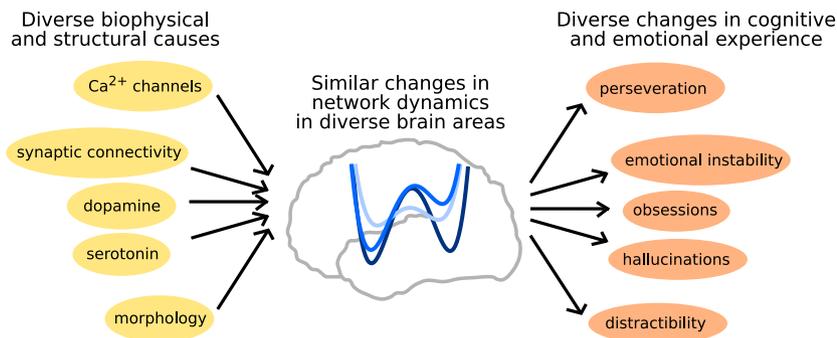
<https://doi.org/10.1016/j.bpsc.2020.01.001>

Mental illnesses are highly complex, temporally dynamic phenomena (1). Variables across a vast range of timescales—from milliseconds to generations—and levels—from subcellular to societal—interact in complex manners to result in the dynamic, rich, and extraordinarily heterogeneous temporal trajectories that are characteristic of the personal and psychiatric histories evident in mental health services across the world. The dynamic and complex nature of these phenomena represents a substantial challenge to our ability to understand mental illnesses and to treat them. The neglect of the temporal aspects of these phenomena may have occurred in part because longitudinal studies have traditionally been more challenging, and hence much research has focused on cross-sectional samples. However, variation observed between individuals will only rarely be informative about individual variation over time (2), and it is arguably the latter variation that matters more in treatment settings. Time, we suggest, matters, and these dynamical aspects can and need to be addressed directly.

When multiple variables interact with one another in a complex manner over time, this interaction gives rise to dynamical systems (DS) that obey certain rules regardless of the particular nature of the variables involved. The behavior of such systems is studied in the mathematical framework of DS theory (DST). DST formalizes the complex interaction of

variables by a set of differential (if formulated in continuous time) or recursive (if in discrete time) equations. It provides a powerful and general mathematical language and toolbox for examining phenomena in such systems that are generic—that is, independent from their specific physical realization—and that exist across timescales. These phenomena include, for instance, oscillations, synchronization among units of a system, attractor states, phase transitions, and deterministic chaos. Although generic and formulated in an abstract language, these phenomena are not merely conceptual or even metaphorical, but “real.” They are experimentally and clinically accessible and quantifiable processes that can be measured and inferred from data, and that determine and predict future developments and prescribe how to best influence the system. As such, they should have a prominent place in guiding interventions.

As we will argue in this article, DST may serve as a kind of hub, a central layer of convergence or level of nervous system description at which phenomena relevant to mental illness could be understood, explained, classified, and predicted. DST represents a layer of convergence in the sense that a number of very different, seemingly unrelated physiological and anatomical processes may give rise to similar alterations in network dynamics and behavior (Figure 1). This may explain why different causal factors and pathogenic routes may give



**Figure 1.** Network dynamics as a layer of convergence. A number of different physiological and structural processes (left) may give rise to similar alterations in network dynamics (center) that, depending on where in the brain they manifest, may give rise to a variety of different cognitive and emotional processes and psychiatric symptoms (right).

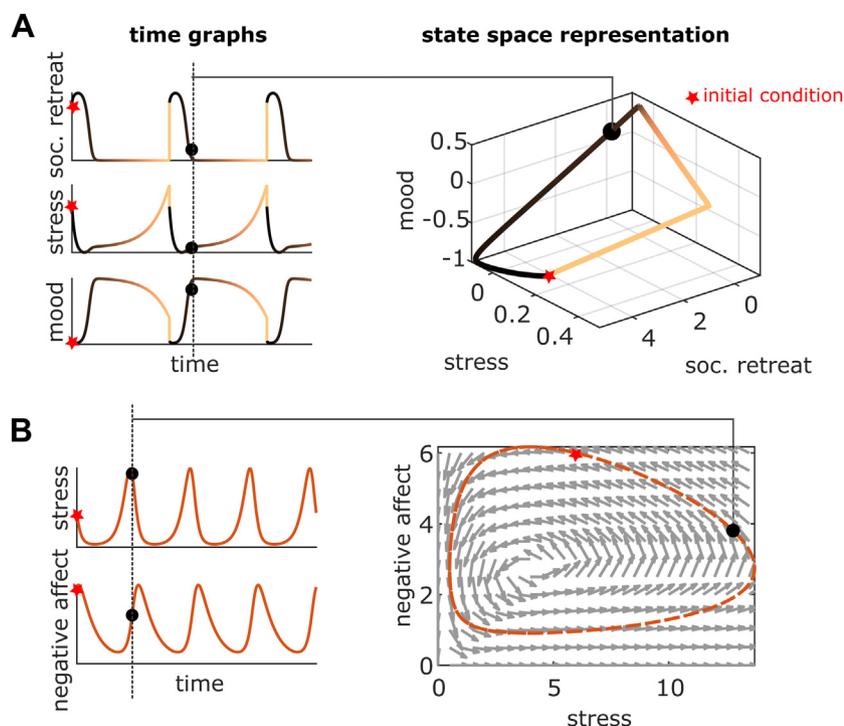
rise to similar phenomena [see (3)]. At the same time, the same changes in network dynamics may be associated with a variety of different symptoms (Figure 1), depending on the brain areas in which these dynamical alterations are mostly expressed. This emphasis on the DS level also bears important implications for the treatment of mental illness, as discussed in Implications: Dynamics as Treatment Targets.

In this article, we introduce important DST concepts (Figure 2) and phenomena directly within the context of neuroscientific and psychiatric observations that they may account for [see also Roberts *et al.* (4), Breakspear (5), and Wang and Krystal (6)], and to illustrate them based on the same formal example of a DS, a recurrent neural network (RNN) model (Figure 3A), with more formal background included in

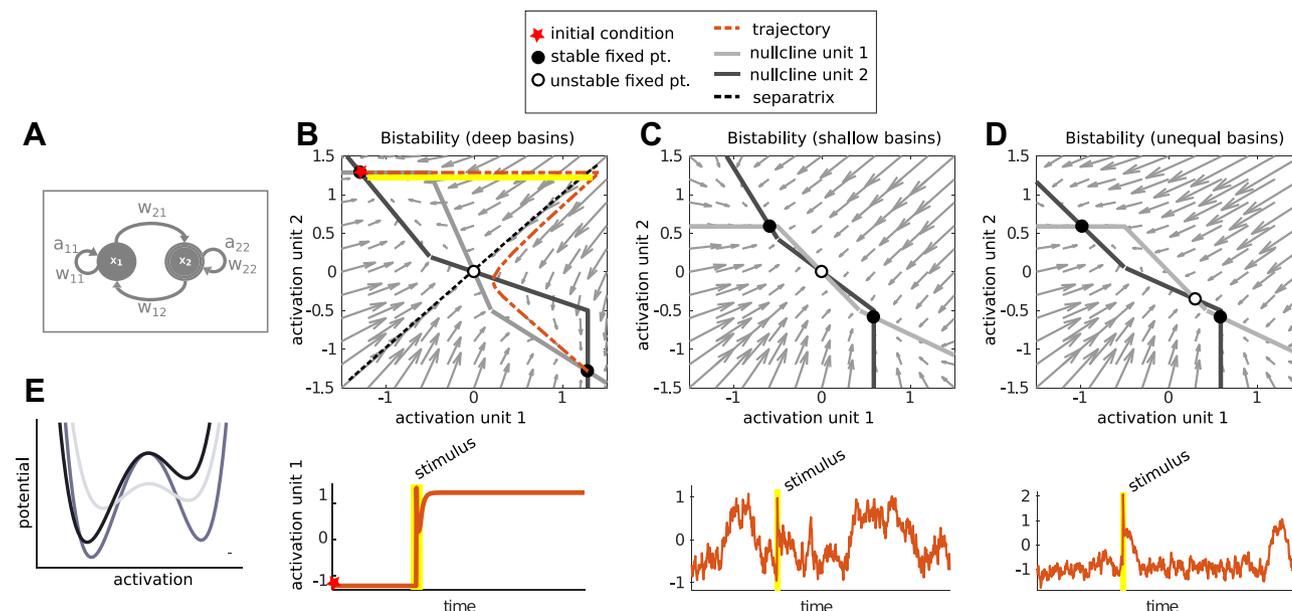
the Supplement. We also briefly address how DS can be inferred from observations.

### DYNAMICAL PHENOMENA AND THEIR POTENTIAL RELATIONSHIP TO PSYCHIATRIC CONDITIONS

A DS is described by a set of system variables (such as membrane potentials or symptom strengths) and equations governing their temporal evolution (see the DST primer in the Supplement). A comprehensive geometrical representation of a DS is its state space, which is the space spanned by all its dynamical variables, as illustrated in Figure 2A. A powerful property of the state space representation is that it provides a complete description of the system's state, behavior, and (in



**Figure 2.** Time graphs and state spaces. **(A)** A central concept in dynamical systems theory is that of a state space (right). A state space is the space spanned by all dynamical variables of a system, which in this psychological example were taken to be mood, stress, and social (soc.) retreat. A trajectory in the state space corresponds to the temporal coevolution of the dynamical variables over time, i.e., there is a 1:1 correspondence between points on the trajectory and the state of all dynamical variables when depicted as a function of time (left). The color coding of the trajectory illustrates time progression. **(B)** Another central concept is that of a flow field (right), where the vectors indicate the direction and magnitude of flow (change) at each point in state space, illustrated here with a 2-dimensional example. Examples were constructed based on the Lotka-Volterra equations.



**Figure 3.** Example of multistability in a recurrent neural network (RNN). **(A)** Structure of a 2-unit “toy” RNN (see equation 6 and parameter values used in the Supplement). **(B)** Flow field for the RNN in **(A)**, with gray arrows marking direction and magnitude of the flow. Gray-shaded lines are the nullclines of unit 1 and unit 2, where the flow of either one of the two system variables vanishes, and solid and open circles show stable and unstable fixed points, respectively. The black dashed line separates the basins of attraction of the two stable fixed points. The red dashed line shows an example of a deterministic trajectory starting from the initial condition indicated by the red star (located on one of the system’s two point attractors), after a brief stimulus (yellow) to unit 1. **(C)** Flow field for the same system as in **(B)**, with network parameters slightly changed, causing the system’s attractors to move closer together and their basins to become shallower. **(D)** Flow field for the same system as in **(B)** and **(C)**, with parameters slightly changed such that the symmetry between attractor states is broken. The system now has one attractor with a steeper basin than the other. The bottom sections of **(B–D)** show the activation in time of unit 1, with the period of stimulation indicated in yellow. In **(C, D)**, noise was added to the system. While in **(B)**, the network maintains unit 1’s high activation by remaining in the high-state attractor, **(C)** shows how activity spontaneously switches between the two attractors because of the presence of noise and shallower attractor basins. In **(D)**, the system remains only briefly in the high-state attractor because of its small basin of attraction, from which it is kicked out by the noise after relatively short dwelling times. **(E)** Schematic potential landscape depicting the extent and depth of the basins of attraction of the systems in **(B)** (dark gray), **(C)** (light gray), and **(D)** (black). Potential minima correspond to the attractor states. MATLAB code for these simulations is available at <https://github.com/DurstewitzLab>.

the deterministic case) future fate: a point in this space exhaustively specifies the system’s current state (i.e., the current values of all variables describing the system), and the so-called flow (vector) field (the arrows in Figure 2B) completely specifies how the system will evolve in time when released at any point in this space (namely, along the direction indicated by the vectors). The temporal evolution of the system’s state within this space when started from a specific initial condition is represented by its trajectory (Figure 2). In essence, the system’s trajectory in state space shows how all variables jointly evolve in time; there is a 1:1 correspondence between such a trajectory and the more familiar time graphs of all variables (Figure 2).

Consider as a very simple psychological example the interaction between psychological stress, mood, and social retreat, as depicted in Figure 2A. As stress levels increase, with some delay mood will decline, which in turn may lead to an increased tendency to retreat from the world and social interactions. As a consequence, stress levels may drop again, mood will tend to increase, and the person may increasingly engage again in social and job-related responsibilities, potentially starting the whole cycle all over again. Such cyclic interactions between variables are commonly observed in the setting of mental health and are important tools, for instance, in

case formulations in psychotherapy. Indeed, interactions between symptoms have been argued to characterize the long-term course of illnesses better than standard latent-factor models (3,7).

### Attractor Dynamics and Multistability

Figure 3B illustrates the flow field for a simple 2-dimensional formal example of a DS, a 2-unit RNN (Figure 3A). The flow field indicates a specific geometry of the state space that determines the fate of trajectories when released at specific initial conditions: in this case, the state space contains 3 fixed points, points at which the flow becomes exactly zero in all directions (i.e., the vectors vanish). Of them, 2 are stable (solid dots) in the sense that activity converges to them from all directions; hence, small displacements (perturbations) decay back to them. Such stable fixed points are also called point attractors, and they are surrounded by a basin of attraction, which is the set of all points from which activity converges to the respective point attractor. The fixed point in the center (open dot), in contrast, is unstable with activity diverging along at least 1 direction (fixed points with both converging and diverging directions are called saddle points). If noise is added to the DS, it may cause trajectories to eventually cross

the energy ridge between attractors (Figure 3E). The likelihood of such transitions or, conversely, the dwell times within specific states, will depend on the noise amplitude and the steepness of the attractor basins, i.e., the magnitude of the opposing flow (Figure 3C, D). This gives rise to a phenomenon called metastability (8), where noise-induced perturbations can cause the system to hop around different attractor states (Figure 3C).

DS may harbor many different stable fixed points or other attractor objects. Such multistability—that is, the coexistence of many attractor states—has been proposed to underlie functions such as working memory (9,10), with each fixed point corresponding to the active maintenance of a different memory item. The idea is that different briefly presented stimuli would push the network into one of the different stimulus-specific attractor states (Figure 3B), which by virtue of their attractor property would maintain an active representation of the stimulus even after its removal. Physiologically, the attractor states may correspond to elevated firing rates (termed “persistent activity”) in the respective subset of stimulus-selective neurons [e.g., (11,12)] and may be established through strong recurrent excitation within this subpopulation (or “cell assembly”) (9,13). Attractor dynamics are thought to play a role also in a variety of other cognitive processes, including decision making (14–18), probabilistic (Bayesian) inference (19), the formation and maintenance of beliefs (20,21), or processes such as memory recollection and pattern completion (19,22–24). Formally, models of reinforcement learning are DS as well (25,26) that may settle into stable fixed points in a stationary environment.

Profound alterations in attractor dynamics may affect mental functions. As a biophysical-level example, dopamine via its synaptic and ionic actions can regulate the width and steepness of basins of attraction (Figure 3E), with the direction of change depending on the receptor subtype ( $D_1$  vs.  $D_2$  class) primarily stimulated (27–33). This could alter the trade-off between cognitive flexibility, which is supported by flat attractor basins that ease moving among representations, and working memory and goal orientation, which are facilitated by deep and wide basins that protect the current focus (27,34). Via these dynamical mechanisms, the changes in the dopaminergic system known in schizophrenia may therefore account for the

observed deficits in both working memory and cognitive flexibility (27,35,36).

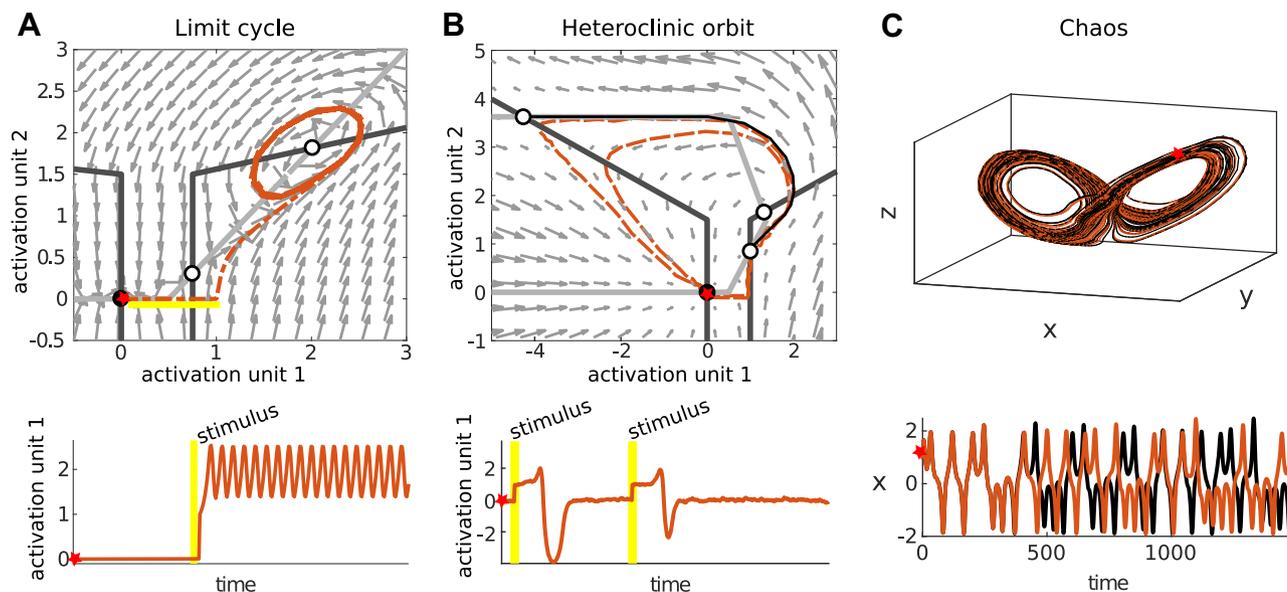
Consider as another example impaired emotion regulation in depression. Ramirez-Mahaluf and Compte (37) viewed this as emerging from the mutually inhibitory interaction between an emotional and a cognitive hub, namely the ventral anterior cingulate cortex and the dorsolateral prefrontal cortex, respectively. According to their model, high glutamatergic tone in the ventral anterior cingulate cortex results in overly stable attractor states that inhibit cognitive activation in the dorsolateral prefrontal cortex. Within a certain parameter regime, this could be counteracted by serotonin-induced hyperpolarization of ventral anterior cingulate cortex neurons through selective serotonin reuptake inhibitors. This idea is illustrated in Figure 3B–D with 2 units (which one may think of as representing 2 network hubs in this context) with strong self-excitation but mutual inhibition (i.e., negative weights  $w_{12}$  and  $w_{21}$  and positive weights  $w_{11}$  and  $w_{22}$  in Figure 3A). As illustrated, by either increasing the amount of self-excitation in 1 of the 2 hubs or through an imbalance in the feedback between the two (Figure 3D), 1 of the 2 attractor basins may strongly expand at the expense of the other. At the psychological level, this type of account would also explain why strong rumination and negative mood (reflecting a strong emotion attractor) concur with a lack of attention and impaired decision making (38,39), or why increased fear may inhibit performance under high cognitive load and vice versa (40).

As in the example of working memory, strong attractor states often arise through positive feedback loops. For instance, stressful life events predict depression but are also caused by depression (41), raising the possibility that after a first life event, further life events may be caused by the depressed state, leading to a mutually reinforcing feedback loop between stress and depression. Conflict states in couples (42) and groups (43) may manifest through similar attractor dynamics, with positive feedback loops leading to escalation, emphasizing the role of de-escalation techniques. While space constraints prevent a more detailed discussion, we point to similar studies highlighting the role of attractor dynamics in the domains of ketamine (32,44), dopamine and schizophrenia (27–32,45), depression (46), attention-deficit/hyperactivity disorder (ADHD) (47–51), obsessive-compulsive disorder (52,53), and posttraumatic stress disorder (54–56) (see also Table 1).

**Table 1. Psychiatric Symptoms and Their Possible Dynamical Systems Interpretation**

Symptoms	Associated Changes in Attractor Dynamics
Perseveration, Dissociation, Obsessions, Compulsions	Overly steep attractor basins
Distractor Susceptibility/Inattentiveness, Associative Hopping, Incoherent and Disorganized Thought, Hallucinations	Overly flat attractor basins or increased noise levels
Deficits in Parametric Working Memory, Jumping to Conclusions (Failure to Integrate Information)	Alterations in line attractor configurations
Rumination and Recurring Chains of Thought, Stereotypical Movement Patterns, Persistence of Invalid Belief Sets	Overly steep limit-cycle attractors or heteroclinic channels
Altered Time Perception, Slowed-Down Mental Processes	Alterations in flow around attractor ghosts
Lucid Moments in Amnesia, Epileptic Seizures, Sudden Transitions Between Disease Stages, Resistance to Therapy	Bifurcations
Increased Variability in Affective States, Disorganized Thought, High Distractibility	Too high chaoticity
Reduced Cognitive Flexibility	Too low chaoticity

## Mental Illnesses as Disorders of Network Dynamics



**Figure 4.** Examples of different sequential phenomena in dynamical systems, illustrated with a recurrent neural network (RNN). Panels depict flow fields (top row) and time graphs of unit 1 (bottom row) for different parameter settings of an RNN (equation 6 in the Supplement). Gray-shaded lines in flow fields mark nullclines of units 1 and 2, red dotted lines show one trajectory starting from an initial condition (red star), and light yellow lines mark the brief presentations of positive external stimuli. **(A)** Bistability among a stable limit cycle surrounding the right unstable fixed point, and a stable fixed point in a 2-unit RNN (Figure 3A). A brief stimulus to unit 1 takes the system from its stable fixed point to the stable limit cycle. Vectors are all normalized to same length for better visualization of flow direction. **(B)** Heteroclinic orbit (shown in black) connecting the system's 2 saddle nodes. In this case, the heteroclinic channel created by this orbit is itself not attracting [in contrast to examples in Rabinovich *et al.* (69)], such that nearby trajectories tend to diverge from it (but note that, in the deterministic case, the system would continue to move on the heteroclinic orbit if placed exactly on it). Yet, this unstable heteroclinic channel still influences the behavior of the system in the sense that brief perturbations through an external stimulus (shown in light yellow) tend to cause trajectories to move slowly in its vicinity (just below it) until they return to the stable fixed point at the bottom. **(C)** The famous chaotic Lorenz attractor (93), reproduced by an RNN (equation 6 in the Supplement) statistically inferred from trajectories drawn from the Lorenz system (140). As characteristic of chaotic systems, two very close-by initial conditions may lead into very different activation patterns in the longer run, as displayed in the bottom panel for one of the RNN variables.

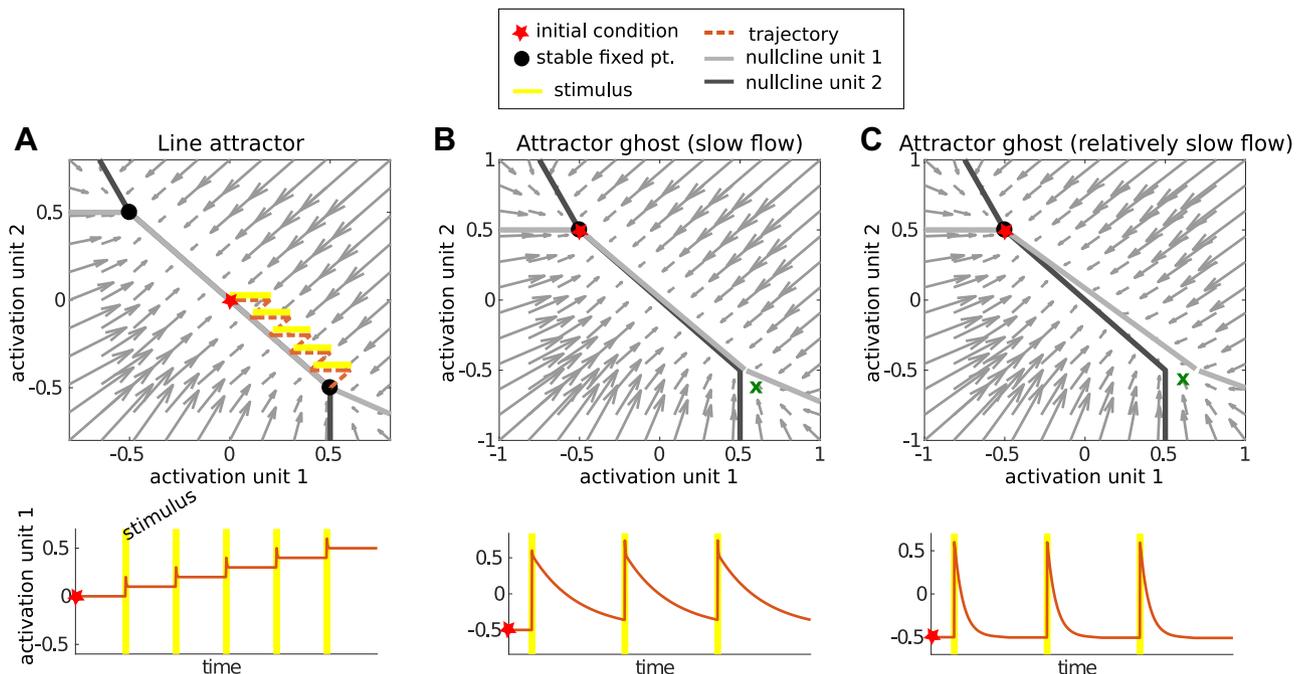
### Sequential Phenomena: Limit Cycles and Heteroclinic Channels

Figure 4A illustrates another setup of the RNN. A slight change in some of the system parameters (see Supplement) gives rise to a different set of phenomena: rather than converging to a stable fixed point, the RNN now continues to periodically oscillate. It may not be a simple harmonic (sinusoidal-type) oscillation, however, but a more complex waveform that is repetitively produced. This complex but still periodic waveform represents another type of attractor state, termed a stable limit cycle (just as with fixed points, there are also unstable limit cycles). Limit cycles can become complicated in appearance, with multiple different minima and maxima and very long periods in which the system's trajectory does not precisely retrace itself, although the cycles always close up eventually. Limit cycles often result from interacting positive and negative feedback loops, which are ubiquitous in the nervous system. They may represent sequences that are to be repetitively produced, such as potentially complex but still relatively stereotypical motor programs or movement patterns (57–60).

Stereotypical, repeating movement patterns are observed in many neurological and some psychiatric conditions (61,62). In general, nonlinear oscillations—the equivalent of limit cycles in the time domain—are a hallmark of nervous system activity (63), and specific alterations, for instance in the gamma or

delta frequency band, have indeed been described in schizophrenia (64) or ADHD (65). At a higher cognitive level, perseveratively recurring chains of the same thoughts may potentially be generated this way. Stable limit cycles may also underlie many types of symptom clusters that emerge in periodic intervals (66–68).

There are also other, more flexible ways to generate sequences in DS, as illustrated in Figure 4B (69,70). Here, orbits connect a chain of saddle points, that is, half-stable fixed points toward which activity converges from some directions but leaves along others. The curves that connect the different saddle points are called heteroclinic orbits (Figure 4B), and the whole arrangement of heteroclinic orbits connecting a chain of saddle points has been termed a heteroclinic channel (HC) (69,70). The HC acts like an attractor, pulling in trajectories from the vicinity that then, with a bit of noise, travel along the curves connecting the saddle points, which may implement a sequence of thoughts or actions. Unlike a limit cycle, the HC is not necessarily automatically repeating—it may start and terminate in a stable fixed point (as in the example in Figure 4B). More importantly, this arrangement is much more flexible: while limit cycles determine a rather rigid sequence of events, in an HC, saddle points could more easily be removed from or added to the already existing sequence through proper parameter changes, making this a more plausible account for



**Figure 5.** Example of line attractors and slow flow in a recurrent neural network (RNN). Panels (A–C) depict flow fields for slightly different parameter settings of a 2-unit RNN (equation 6 and parameters used in Supplement). Panel (A) shows a line attractor. Gray-shaded lines mark the nullclines of units 1 and 2. Red dotted lines show one trajectory starting from its initial condition (red star) and briefly pushed away from the line attractor by stimulus pulses (indicated in light yellow), indicating that the line attractor integrates stimuli. (B) When the parameters of this line attractor configuration are changed, the system’s bottom-right fixed point disappears and leaves behind an attractor ghost. Near this attractor ghost, the flow is very slow or (C) relatively slow, depending on how far the system’s parameters were moved away from the truly attracting configuration. The bottom panels show the activation of unit 1 for systems in (A–C), respectively, starting from the initial condition (red star), and stimulated repeatedly as indicated by the yellow lines. Note that for clarity we omitted trajectories and stimuli from (B) and (C). The stimuli in (B) and (C) take the state of the system to the spots in state space indicated by the green crosses.

higher cognitive functions (e.g., syntactical sequences) than limit cycles (69). Indeed, it has been suggested that belief sets evolve as HCs in the course of psychotherapy (71).

Whether psychiatric phenomena with periodic behavior, e.g., alternation between relapse and remission episodes, can be better described in terms of limit cycles, HCs, or hopping among metastable states, is a difficult but, in principle, empirically tractable question. Differentiating among these scenarios could have important implications for optimal treatment, in terms of both the type and the timing of an intervention (72).

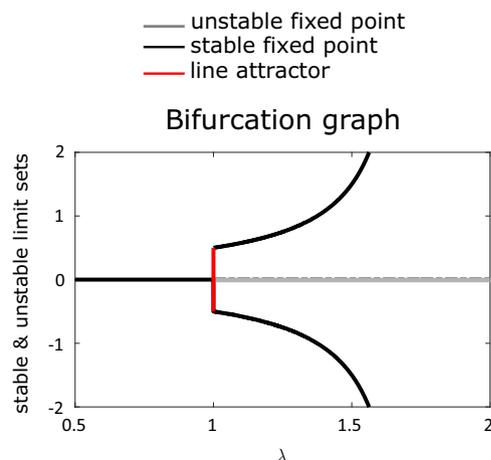
### Attractor Ghosts and the Regulation of Flow

Another important phenomenon in DS is that of attractor ruins (73,74), also known as attractor ghosts (75), which are quasi- or semi-attracting states (76). These are attractors that are almost stable, i.e., to which trajectories still converge along most directions but may slowly escape along others (Figure 5B, C). In these scenarios, the system’s parameters are very close to a configuration that would yield a true attractor, just not quite there (Figure 6).

This phenomenon comes with important and interesting implications that differentiate these objects from either true attractors or clearly unstable objects. Imagine a scenario where attractor valleys (Figure 3E) become perfectly flat along one or more directions. This gives rise to a so-called line attractor

where the fixed points form a line, ring, or plane (77–80), a continuum of neutrally stable fixed points along which there is neither convergence nor divergence (Figure 5A). Line attractors have been proposed to underlie phenomena such as parametric working memory (78) where a continuously valued quantity [like a flutter frequency (81) or spatial position (82)] has to be retained in working memory. An attractor ruin results, for instance, if we now slightly “detune” the line attractor (Figure 5B). This leads to new effective time constants that are largely independent from the system’s intrinsic (e.g., biophysical) time constants, a phenomenon that has been exploited for interval timing in neural systems (77,83). Too-wide detuning may in turn account for timing problems, specifically a speed-up of the internal clock, as evident in Parkinson’s disease (84) or ADHD (85), given that the dopaminergic system has been linked to alterations in (interval) time perception and production (86,87). Too-narrow tuning, on the other hand, could produce a slowing down of time perception, as in patients with bipolar disorder (88).

Hence, trajectories considerably slow down and tend to prevail in attractor ruins. As with HC, this phenomenon could also be exploited for flexible sequence generation with trajectories traveling among attractor ruins (89). In consequence, alterations in attractor ruins may cause characteristic symptoms, e.g., slowed-down mental processing as often observed in patients with major depressive disorder (90,91).



**Figure 6.** Bifurcations. Example of a bifurcation graph for the system in Figure 5 with fixed points plotted as a function of network parameter  $\lambda$ . Here,  $\lambda$  is a factor that regulates the strength of the units' self-excitation. With  $\lambda < 1$ , the network exhibits only a single stable fixed point, and then switches to a line attractor for  $\lambda = 1$  (as in Figure 5A), and finally harbors 2 stable and 1 unstable fixed points for  $\lambda > 1$  (as in the systems in Figure 3).

## Chaos

Chaos is a strange phenomenon where a deterministic DS exhibits aperiodic and irregular behavior even in the absence of noise, with the system's states never quite repeating themselves (Figure 4C) (75,92). The state of chaos can still be an attractor, pulling trajectories from a larger basin of attraction into a bounded region of state space within which trajectories would continue to travel forever yet would not form a closed orbit (i.e., limit cycle) (75). Unlike fixed-point and limit-cycle attractors, chaotic attractors have at least one direction along which trajectories diverge, yet get reinjected into the same volume of state space (75). Because of this divergence, activity on the chaotic attractor is highly (exponentially) sensitive to perturbations and minimal differences in initial conditions (Figure 4C, bottom)—the famous butterfly effect (93).

The fact that activity in chaotic attractors is irregular yet not random, retaining a certain sequential structure, may also be beneficial for certain cognitive and coding purposes (94). In a sense, it creates deterministic variation around a central theme, which may be relevant to cognitive search and creativity (95,96). Especially interesting from a computational perspective is the phenomenon of chaotic itinerancy (73,74), where trajectories chaotically traverse between different attractor ruins (see Attractor Ghosts and the Regulation of Flow), a setup that has been exploited for dynamic and flexible sequence production and recognition (89).

Somewhat surprisingly, placing neural systems at the edge of chaos, or slightly within a chaotic regime (97–100), has important computational benefits: here, the system naturally expresses complex temporal structure while at the same time hanging on to external stimulus information for a while. In contrast, if the system is too regular (too convergent), it exhibits no interesting internal behavior, while if it is too chaotic (too divergent), it quickly forgets about external stimuli. Consequently, if the brain leaves this computationally optimal regime and migrates either too much into the regular or too much into

the chaotic range, problems may ensue. Indeed, patients with posttraumatic stress disorder show a greatly reduced heart rate variability (i.e., larger regularity), which is assumed to be indicative of a reduced ability to flexibly respond to incoming information (101,102). Diminished variability in mental states has also been described in older age (103). On the other hand, a highly chaotic regime with its sensitivity to perturbations may account for attentional problems and a high distractibility by external stimuli, e.g., as observed in ADHD (49,104).

As another example, some authors have argued that the seemingly random patterns of thought observed in schizophrenia, which are reflected in associative hopping and disorganized cognition, may be rooted in too-chaotic system dynamics [see, e.g., Paulus and Braff (105)]. In line with this idea, a number of studies observed signatures of increased chaoticity in patients with schizophrenia in electrophysiological and electrodermal recordings (106–109). Mood variations in bipolar disorder have also been characterized as increasingly chaotic patterns (110,111), which are potentially driven by stronger interactions among negative affective states (112); in general, increased coupling among network elements can lead to a chaotic regime (113).

## Phase Transitions and Bifurcations

We have repeatedly highlighted that many dynamical phenomena may be obtained within the very same system (Figures 3–5) just by changing some of its parameters. This phenomenon gives rise to another highly important observation: as system parameters are smoothly changed, we may encounter dramatic and abrupt, qualitative changes in the system's behavior at some critical point (Figure 6). These are points in parameter space, called bifurcation points, where the set of dynamical objects or their properties change, i.e., where certain fixed points, limit cycles, or chaotic objects may come into existence, vanish, or change stability.

This observation in DS is likely to have profound implications for our understanding of crucial transitions, sudden onsets or offsets, or different distinct phases in psychiatric illnesses. That neural systems may undergo critical bifurcations with dramatic consequences is comparatively well established in epilepsy (114,115), where one has relatively clear electrophysiological signatures that allow different types of bifurcations to be identified and distinguished [see also (94,114,116,117,118,119,120)].

At a more cognitive level, bifurcations may account for sudden transitions observed in behavioral choices and the accompanying neural activity during the learning of a new rule (121). Also, both brief amnesic periods (122), during which stored memories cannot be recalled, as well as so-called lucid moments in dementia (123), where suddenly mnemonic details can be recovered again, suggest that neural systems may sometimes hover at the edge of a bifurcation. Bifurcations may also help to explain why psychopharmacological treatment sometimes helps and in other instances completely fails: Ramirez-Mahaluf *et al.* (124), for instance, mimicked the effects of increased glutamate reuptake and selective serotonin reuptake inhibitors on network activity, and found that while within a certain range "healthy" attractor dynamics could be pharmacologically restored, network changes appeared to be

irreversible by pharmacological means once critical bifurcation points were passed. In such cases, to kick a neural system out of particularly deep attractor states, more profound perturbations (potentially provided by interventions such as electroconvulsive therapy or deep-brain stimulation) may be required (125,126). Profound changes reminiscent of crossing bifurcation points have also been proposed as explanation for therapy resistance in schizophrenia (72). The change between states of depression and health also shows signatures that are typical of a phase transition, namely so-called critical slowing down (similar to that in Figure 5B), where the autocorrelation length of different emotions increases (127).

Hence, from a DS perspective, one may see therapeutic efforts in psychiatry as attempts to prevent certain bifurcations from happening or to induce others.

### Inferring DST Phenomena From Empirical Observations

DST will of course be useful to the clinic only if the discussed phenomena can be measured and characterized. Correlation- or coherence-based analyses (128–130), power spectra (131–133), or tools such as dynamic causal modeling (134–137) have been used for some time to characterize changes in functional and effective connectivity among brain nodes and other temporal properties of neurophysiological time series. However, almost all of these tools are linear or, like hidden Markov models (138,139), come with strong assumptions and restrictions. Linear models cannot produce most of the DS phenomena discussed here, except for simple phenomena such as isolated fixed points, line attractors, or simple (unstable/ neutrally stable) harmonic oscillations. They are therefore not suitable for addressing DS phenomena more generally [see, e.g., (140)]. Some nonlinear aspects of the dynamics can be inferred from scaling laws [(141), but see (142)], perturbation approaches (143,144), change points (121,145), delay embeddings (146), or other properties of the observed time series (23,147). While such methods provide important signatures of specific phenomena (e.g., a bifurcation), they do not return a full picture of the system dynamics.

More recently, however, progress in machine learning made it possible to extract attractor dynamics directly from empirical time series, such as electroencephalogram or functional magnetic resonance imaging measurements (140) or ecological momentary assessments (148), using generic nonlinear DS formulations set up to approximate whatever set of unknown governing equations may have generated the empirical observations (140,149–151). RNNs are particularly well suited for this purpose: there are mathematical theorems that assure us that RNNs are dynamically universal in the sense that (almost) any other DS can be reformulated as a dynamically equivalent RNN that will produce the same flow field and thus dynamics in state space (151–153). Coupled with sophisticated statistical inference and deep-learning methods (140,149,154), RNNs can be trained to reproduce and forecast experimental time series and ultimately to recover the underlying DS itself (140,148). For a more in-depth discussion of these new developments, other methods (146,155,156), and some of the current limitations and caveats, see the Supplement and Koppe *et al.* (140).

### IMPLICATIONS: DYNAMICS AS TREATMENT TARGETS

Computational system dynamics provides an inherently translational language that could be used to describe diverse phenomena at biophysical, cognitive, and even societal levels in the very same DS terms, in the language of state spaces, trajectories, and attractors (69,77,116,157). It thereby enables, for instance, findings in animals to be directly related to findings in humans, or mechanistic DS insights to be transferred from one physiological or behavioral domain to another. Approaches for reconstructing DS from data (140,149,158–161) are even relatively agnostic to the precise measurement modality (except for limitations from a method's temporal or spatial resolution); that is, the same DS may be inferred from neuroimaging, surface electrode, multiple single-unit recordings, or behavioral data.

The most critical contribution of DST is likely an appreciation of dynamical rather than static features as potential targets of interventions. Consider an example from schizophrenia. Traditionally, the focus has been to use medication to directly reverse known physiological aberrations, e.g., through dopaminergic antagonists. However, DS are complex, and many diverse ionic effects may act synergistically to establish a certain dynamical regime (28,162). Restoring only part of the ionic functions underlying the original deficits, e.g., only gamma-aminobutyric acidergic transmission, could—counterintuitively—make the situation even worse (9). On the other hand, from a DS perspective, pharmacological agents may not have to target exactly those transmitter systems that are most compromised in a disease. Perhaps it is easier, cheaper, or more biocompatible to use compounds that target alternative mechanisms, for instance, cellular calcium channels, with exactly the same implications for dynamics that dopaminergic drugs would have. Because there are usually many different and mutually redundant routes to the same dynamical phenomena, very different treatments could have similar effects. Moreover, some dynamic feedback loops in the brain may be much more sensitive to parameter changes than others, rendering them more effective targets for treatment than others. Appreciation of dynamical properties may hence open up new paths for intervention.

One other implication is that assessing and monitoring the system dynamics in patients or at-risk subjects may be more informative than the current approach of examining subjective phenomena by asking individuals to average over periods ranging from weeks to months (thus averaging out temporal dynamics). The DST perspective may also help us to understand how seemingly unrelated phenomena are truly connected at a deeper level (leading, e.g., into comorbidities): perhaps the brain becomes generally vulnerable to a specific type of alteration of its dynamical regimes (e.g., due to a transmitter imbalance); depending on which brain areas are affected most by these dynamical alterations, they will find their incarnation in different bundles of symptoms (see Figure 1). For instance, while hyperstable attractor states in auditory areas may cause tinnitus, the same alterations in the orbitofrontal cortex may be associated with perseveration of suboptimal responses.

In conclusion, appreciating the dynamical properties of mental illnesses could have profound implications for how we

## Mental Illnesses as Disorders of Network Dynamics

diagnose, classify, predict, and treat psychiatric symptoms. Currently, however, this field is still very much in its infancy, and much more systematic empirical studies that directly address DST mechanisms in psychiatry are needed, potentially building on new methodological developments in the fields of machine and deep learning (see Inferring DST Phenomena From Empirical Observations).

## ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by the German Research Foundation (DFG) (Grant Nos. Du 354/8-2 and Du 354/10-1 [to DD] and Grant Nos. TRR265: A06 [to DD, GK] and TRR265: B08 [to GK]) and from the German Federal Ministry of Education and Research (BMBF) within the e:Med program (Grant Nos. 01ZX1311A [SP7] and 01ZX1314G [SP10] [to DD]). Dr. Huys acknowledges support from the University College London Hospitals, National Institute for Health Research, Biomedical Research Centre.

A version of this article was published as a preprint on arXiv: 1809.06303.

The authors report no biomedical financial interests or potential conflicts of interest.

## ARTICLE INFORMATION

From the Department of Theoretical Neuroscience (DD, GK) and Department of Psychiatry and Psychotherapy (GK), Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim; and Faculty of Physics and Astronomy (DD), Heidelberg University, Heidelberg, Germany; Division of Psychiatry and Max Planck UCL Centre for Computational Psychiatry and Ageing Research (QJMH), University College London, London, United Kingdom; and Translational Neuromodeling Unit (QJMH), Institute for Biomedical Engineering, University of Zurich and ETH Zurich, Switzerland.

Address correspondence to Daniel Durstewitz, Ph.D., Department of Theoretical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, J5, 68158 Mannheim, Germany; E-mail: [daniel.durstewitz@zi-mannheim.de](mailto:daniel.durstewitz@zi-mannheim.de); or Georgia Koppe, Ph.D., Department of Theoretical Neuroscience and Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, J5, 68158 Mannheim, Germany; E-mail: [georgia.koppe@zi-mannheim.de](mailto:georgia.koppe@zi-mannheim.de).

Received Dec 17, 2019; accepted Jan 6, 2020.

Supplementary material cited in this article is available online at <https://doi.org/10.1016/j.bpsc.2020.01.001>.

## REFERENCES

- Bystritsky A, Nierenberg A, Feusner J, Rabinovich M (2012): Computational non-linear dynamical psychiatry: A new methodological paradigm for diagnosis and course of illness. *J Psychiatr Res* 46:428–435.
- Molenaar PC, Campbell CG (2009): The new person-specific paradigm in psychology. *Curr Dir Psychol Sci* 18:112–117.
- Kendler KS, Zachar P, Craver C (2011): What kinds of things are psychiatric disorders? *Psychol Med* 41:1143–1150.
- Roberts JA, Friston KJ, Breakspear M (2017): Clinical applications of stochastic dynamic models of the brain, Part II: A Review. *Biol Psychiatry Cogn Neurosci Neuroimaging* 2:225–234.
- Breakspear M (2017): Dynamic models of large-scale brain activity. *Nat Neurosci* 20:340–352.
- Wang XJ, Krystal JH (2014): Computational psychiatry. *Neuron* 84:638–654.
- Borsboom D, Cramer AO, Schmittmann VD, Epskamp S, Waldorp LJ (2011): The small world of psychopathology. *PLoS One* 6:e27407.
- Balaguer-Ballester E, Moreno-Bote R, Deco G, Durstewitz D (2017): Metastable dynamics of neural ensembles. *Front Syst Neurosci* 11:99.
- Durstewitz D, Seamans JK, Sejnowski TJ (2000): Neurocomputational models of working memory. *Nat Neurosci* 3:1184–1191.
- Wang XJ (2001): Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci* 24:455–463.
- Funahashi S, Bruce CJ, Goldman-Rakic PS (1989): Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol* 61:331–349.
- Fuster J (2015): *The Prefrontal Cortex*, 5th ed. London, UK: Academic Press.
- Hebb DO (1949): *The Organization of Behavior*. New York: Wiley.
- Wang X-J (2008): Decision making in recurrent neuronal circuits. *Neuron* 60:215–234.
- Wang X-J (2002): Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36:955–968.
- Albantakis L, Deco G (2009): The encoding of alternatives in multiple-choice decision-making. *BMC Neuroscience* 10(suppl 1):166.
- Ratcliff R, McKoon G (2008): The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput* 20:873–922.
- Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD (2006): The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev* 113:700–765.
- Lengyel M, Kwag J, Paulsen O, Dayan P (2005): Matching storage and recall: Hippocampal spike timing-dependent plasticity and phase response curves. *Nat Neurosci* 8(12):1677–1683.
- Heskes T (2002): Stable fixed points of loopy belief propagation are local minima of the Bethe free energy. *Adv Neural Inf Process Syst* 15:359–366.
- Adams RA, Napier G, Roiser JP, Mathys C, Gilleen J (2018): Attractor-like dynamics in belief updating in schizophrenia. *J Neurosci* 38:9471–9485.
- Hopfield JJ (1982): Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A* 79:2554–2558.
- Wills TJ, Lever C, Cacucci F, Burgess N, O'Keefe J (2005): Attractor dynamics in the hippocampal representation of the local environment. *Science* 308:873–876.
- Ratcliff R (1978): A theory of memory retrieval. *Psychol Rev* 85:59–108.
- Durstewitz D (2017): *Advanced Data Analysis in Neuroscience: Integrating Statistical and Computational Models*. Cham, Switzerland: Springer.
- Gershman SJ (2015): A unifying probabilistic view of associative learning. *PLoS Comput Biol* 11:e1004567.
- Durstewitz D, Seamans JK (2008): The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-o-methyltransferase genotypes and schizophrenia. *Biol Psychiatry* 64:739–749.
- Durstewitz D, Seamans JK, Sejnowski TJ (2000): Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *J Neurophysiol* 83:1733–1750.
- Lapish CC, Balaguer-Ballester E, Seamans JK, Phillips AG, Durstewitz D (2015): Amphetamine exerts dose-dependent changes in prefrontal cortex attractor dynamics during working memory. *J Neurosci* 35:10172–10187.
- Gruber AJ, Solla SA, Surmeier DJ, Houk JC (2003): Modulation of striatal single units by expected reward: A spiny neuron model displaying dopamine-induced bistability. *J Neurophysiol* 90:1095–1114.
- Maia TV, Cano-Colino M (2015): The role of serotonin in orbitofrontal function and obsessive-compulsive disorder. *Clin Psychol Sci* 3:460–482.
- Murray JD, Anticevic A, Gancsos M, Ichinose M, Corlett PR, Krystal JH, et al. (2014): Linking microcircuit dysfunction to cognitive impairment: Effects of disinhibition associated with schizophrenia in a cortical working memory model. *Cereb Cortex* 24:859–872.
- King R, Barchas JD, Huberman B, editors. *Theoretical Psychopathology: An Application of Dynamical Systems Theory to Human Behavior*. Synergetics of the Brain; 1983 1983//; Berlin, Heidelberg: Springer Berlin Heidelberg.
- Ueltzhöffer K, Armbruster-Genç DJ, Fiebach CJ (2015): Stochastic dynamics underlying cognitive stability and flexibility. *PLoS Comput Biol* 11:e1004331.

35. Ambruster DJ, Ueltzhöffer K, Basten U, Fiebach CJ (2012): Prefrontal cortical mechanisms underlying individual differences in cognitive flexibility and stability. *J Cogn Neurosci* 24:2385–2399.
36. Floresco SB, Block AE, Maric T (2008): Inactivation of the medial prefrontal cortex of the rat impairs strategy set-shifting, but not reversal learning, using a novel, automated procedure. *Behav Brain Res* 190:85–96.
37. Ramirez-Mahaluf JP, Compte A (2018): Serotonergic modulation of cognition in prefrontal cortical circuits in major depression. In: Anticevic A, Murray JD, editors. *Computational Psychiatry: Mathematical Modeling of Mental Illness*. London, UK: Academic Press, 27–46.
38. Gotlib IH, Joormann J (2010): Cognition and depression: Current status and future directions. *Annu Rev Clin Psychol* 6:285–312.
39. Lyubomirsky S, Kasri F, Zehm K (2003): Dysphoric rumination impairs concentration on academic tasks. *Cognit Ther Res* 27:309–330.
40. Vytal K, Cornwell B, Arkin N, Grillon C (2012): Describing the interplay between anxiety and cognition: From impaired performance under low cognitive load to reduced anxiety under high load. *Psychophysiology* 49:842–852.
41. Kendler KS, Karkowski LM, Prescott CA (1999): Causal relationship between stressful life events and the onset of major depression. *Am J Psychiatry* 156:837–841.
42. Gottman J, Swanson C, Murray J (1999): The mathematics of marital conflict: Dynamic mathematical nonlinear modeling of newlywed marital interaction. *J Fam Psychol* 13:3–19.
43. Coleman PT, Vallacher RR, Nowak A, Bui-Wrzosinska L (2007): Intractable conflict as an attractor: A dynamical systems approach to conflict escalation and intractability. *Am Behav Sci* 50:1454–1475.
44. Starc M, Murray JD, Santamauro N, Savic A, Diehl C, Cho YT, et al. (2017): Schizophrenia is associated with a pattern of spatial working memory deficits consistent with cortical disinhibition. *Schizophr Res* 181:107–116.
45. Braun U, Harneit A, Pergola G, Menara T, Schaefer A, Betzel RF, et al. (2019): Brain state stability during working memory is explained by network control theory, modulated by dopamine D1/D2 receptor function, and diminished in schizophrenia [published online ahead of print Jun 23]. *bioRxiv*.
46. Cramer AO, van Borkulo CD, Giltay EJ, van der Maas HL, Kendler KS, Scheffer M, et al. (2016): Major depression as a complex dynamic system. *PLoS One* 11:e0167490.
47. Forster S, Lavie N (2016): Establishing the attention-distractibility trait. *Psychol Sci* 27:203–212.
48. Wilens TE, Faraone SV, Biederman J (2004): Attention-deficit/hyperactivity disorder in adults. *JAMA* 292:619–623.
49. Bubl E, Dorr M, Riedel A, Ebert D, Philipsen A, Bach M, et al. (2015): Elevated background noise in adult attention deficit hyperactivity disorder is associated with inattention. *PLoS One* 10:e0118271.
50. Cortese S, Kelly C, Chabernaud C, Proal E, Di Martino A, Milham MP, et al. (2012): Toward systems neuroscience of ADHD: A meta-analysis of 55 fMRI studies. *Am J Psychiatry* 169(10):1038–1055.
51. Hauser TU, Fiore VG, Moutoussis M, Dolan RJ (2016): Computational psychiatry of ADHD: Neural Gain impairments across Marrian levels of analysis. *Trends Neurosci* 39:63–73.
52. Rolls ET, Loh M, Deco G (2008): An attractor hypothesis of obsessive-compulsive disorder. *Eur J Neurosci* 28:782–793.
53. Rabinovich MI, Muezzinoglu MK, Strigo I, Bystritsky A (2010): Dynamical principles of emotion-cognition interaction: Mathematical images of mental disorders. *PLoS One* 5:e12547.
54. Lanius RA, Brand B, Vermetten E, Frewen PA, Spiegel D (2012): The dissociative subtype of posttraumatic stress disorder: Rationale, clinical and neurobiological evidence, and implications. *Depress Anxiety* 29:701–708.
55. Lanius RA, Vermetten E, Loewenstein RJ, Brand B, Schmahl C, Bremner JD, et al. (2010): Emotion modulation in PTSD: Clinical and neurobiological evidence for a dissociative subtype. *Am J Psychiatry* 167:640–647.
56. Sack M, Cillien M, Hopper JW (2012): Acute dissociation and cardiac reactivity to script-driven imagery in trauma-related disorders. *Eur J Psychotraumatol* 3:17419.
57. Kato S, Kaplan HS, Schrodel T, Skora S, Lindsay TH, Yemini E, et al. (2015): Global brain dynamics embed the motor command sequence of *Caenorhabditis elegans*. *Cell* 163:656–669.
58. Marder E, Bucher D (2001): Central pattern generators and the control of rhythmic movements. *Curr Biol* 11:R986–R996.
59. Marder E, Goeritz ML, Otopalik AG (2015): Robust circuit rhythms in small circuits arise from variable circuit components and mechanisms. *Curr Opin Neurobiol* 31:156–163.
60. Russo AA, Bittner SR, Perkins SM, Seely JS, London BM, Lara AH, et al. (2018): Motor cortex embeds muscle-like commands in an untangled population response. *Neuron* 97:953–966. . e8.
61. Ridley R (1994): The psychology of perseverative and stereotyped behaviour. *Prog Neurobiol* 44:221–231.
62. Turner M (1999): Annotation: Repetitive behaviour in autism: A review of psychological research. *J Child Psychol Psychiatry* 40:839–849.
63. Buzsáki G, Draguhn A (2004): Neuronal oscillations in cortical networks. *Science* 304:1926–1929.
64. Uhlhaas PJ, Haenschel C, Nikolic D, Singer W (2008): The role of oscillations and synchrony in cortical networks and their putative relevance for the pathophysiology of schizophrenia. *Schizophr Bull* 34:927–943.
65. Demanuele C, James C, Capilla A, Sonuga-Barke E (2008): Extracting event-related field components through space-time ICA: A study of MEG recordings from children with ADHD and controls. In: Vander Sloten J, Verdonck P, Nyssem M, Haeuelsen J, editors. *4th European Conference of the International Federation for Medical and Biological Engineering*. Berlin, Germany: Springer, 38–42.
66. Chang S-S, Chou T (2018): A dynamical bifurcation model of bipolar disorder based on learned expectation and asymmetry in mood sensitivity. *Comput Psychiatr* 2:205–222.
67. Eldar E, Rutledge RB, Dolan RJ, Niv Y (2016): Mood as representation of momentum. *Trends Cogn Sci* 20:15–24.
68. Eldar E, Niv Y (2015): Interaction between emotional state and learning underlies mood instability. *Nat Commun* 6:6149.
69. Rabinovich MI, Huerta R, Varona P, Afraimovich VS (2008): Transient cognitive dynamics, metastability, and decision making. *PLoS Comput Biol* 4:e1000072.
70. Rabinovich MI, Varona P, Selverston AI, Abarbanel HD (2006): Dynamical principles in neuroscience. *Rev Mod Phys* 78:1213.
71. Kronemyer D, Bystritsky A (2014): A non-linear dynamical approach to belief revision in cognitive behavioral therapy. *Front Comput Neurosci* 8:55.
72. Krystal JH, Anticevic A, Murray JD, Glahn D, Driesen N, Yang G, et al. (2016): Clinical heterogeneity arising from categorical and dimensional features of the neurobiology of psychiatric diagnoses. In: Redish AD, Gordon JA, editors. *Computational Psychiatry*. Cambridge, MA: MIT Press, 293–316.
73. Tsuda I (2001): Toward an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *Behav Brain Sci* 24:793–810.
74. Tsuda I (2015): Chaotic itinerancy and its roles in cognitive neuro-dynamics. *Curr Opin Neurobiol* 31:67–71.
75. Strogatz SH (2018): *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Boca Raton, FL: CRC Press.
76. Balaguer-Ballester E, Lapiš CC, Seamans JK, Durstewitz D (2011): Attracting dynamics of frontal cortex ensembles during memory-guided decision-making. *PLoS Comput Biol* 7:e1002057.
77. Durstewitz D (2003): Self-organizing neural integrator predicts interval times through climbing activity. *J Neurosci* 23:5342–5353.
78. Machens CK, Romo R, Brody CD (2005): Flexible control of mutual inhibition: A neural model of two-interval discrimination. *Science* 307:1121–1124.
79. Seung HS (1996): How the brain keeps the eyes still. *Proc Natl Acad Sci U S A* 93:13339–13344.

## Mental Illnesses as Disorders of Network Dynamics

80. Seung HS, Lee DD, Reis BY, Tank DW (2000): Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron* 26:259–271.
81. Romo R, Brody CD, Hernández A, Lemus L (1999): Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* 399:470–473.
82. Zhang K (1996): Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *J Neurosci* 16:2112–2126.
83. Durstewitz D (2004): Neural representation of interval time. *Neuroreport* 15:745–749.
84. Rammesayer T, Classen W (1997): Impaired temporal discrimination in Parkinson's disease: Temporal processing of brief durations as an indicator of degeneration of dopaminergic neurons in the basal ganglia. *International Journal of Neuroscience* 91:45–55.
85. Rubia K, Halari R, Christakou A, Taylor E (2009): Impulsiveness as a timing disturbance: Neurocognitive abnormalities in attention-deficit hyperactivity disorder during temporal processes and normalization with methylphenidate. *Philos Trans R Soc Lond B Biol Sci* 364:1919–1931.
86. Hass J, Durstewitz D (2014): Neurocomputational models of time perception. In: Merchant H, de Lafuente V, editors. *Neurobiology of Interval Timing*. New York, NY: Springer, 49–71.
87. Rammesayer TH (1993): On dopaminergic modulation of temporal information processing. *Biol Psychol* 36:209–222.
88. Northoff G, Magioncalda P, Martino M, Lee HC, Tseng YC, Lane T (2018): Too fast or too slow? Time and neuronal variability in bipolar disorder—a combined theoretical and empirical investigation. *Schizophr Bull* 44:54–64.
89. Russo E, Treves A (2012): Cortical free-association dynamics: Distinct phases of a latching network. *Phys Rev E Stat Nonlin Soft Matter Phys* 85:051920.
90. Tsourtos G, Thompson J, Stough C (2002): Evidence of an early information processing speed deficit in unipolar major depression. *Psychol Med* 32:259–265.
91. Marazziti D, Consoli G, Picchetti M, Carlini M, Faravelli L (2010): Cognitive impairment in major depression. *Eur J Pharmacol* 626:83–86.
92. Ott E (2002): *Chaos in Dynamical Systems*. Cambridge, UK: Cambridge University Press.
93. Lorenz EN (1963): Deterministic nonperiodic flow. *J Atmos Sci* 20:130–141.
94. Durstewitz D, Gabriel T (2007): Dynamical basis of irregular spiking in NMDA-driven prefrontal cortex neurons. *Cereb Cortex* 17:894–908.
95. Zausner T (1996): The creative chaos: Speculations on the connection between non-linear dynamics and the creative process. *Nonlinear dynamics in human behavior: World Scientific; 1996*. p. 343–349.
96. Schulberg D (1999): Chaos theory and creativity. In: Runco M, Pritzker S, editors. *Encyclopedia of Creativity*, 1st ed. San Diego, CA: Academic Press, 259–272.
97. Bertschinger N, Natschläger T (2004): Real-time computation at the edge of chaos in recurrent neural networks. *Neural Comput* 16:1413–1436.
98. Jaeger H, Haas H (2004): Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* 304:78–80.
99. Legenstein R, Maass W (2007): Edge of chaos and prediction of computational performance for neural circuit models. *Neural Netw* 20:323–334.
100. Sussillo D, Abbott LF (2009): Generating coherent patterns of activity from chaotic neural networks. *Neuron* 63:544–557.
101. Thayer JF, Lane RD (2000): A model of neurovisceral integration in emotion regulation and dysregulation. *J Affect Disord* 61:201–216.
102. Thayer JF, Ahs F, Fredrikson M, Sollers JJ 3rd, Wager TD (2012): A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. *Neurosci Biobehav Rev* 36:747–756.
103. Battaglia D, Thomas B, Hansen EC, Chetouf S, Daffertshofer A, McIntosh AR, *et al.* (2017): Functional connectivity dynamics of the resting state across the human adult lifespan [published online ahead of print Feb 24]. *bioRxiv*.
104. Uebel H, Albrecht B, Asherson P, Börger NA, Butler L, Chen W, *et al.* (2010): Performance variability, impulsivity errors and the impact of incentives as gender-independent endophenotypes for ADHD. *J Child Psychol Psychiatry* 51:210–218.
105. Paulus MP, Braff DL (2003): Chaos and schizophrenia: Does the method fit the madness? *Biol Psychiatry* 53:3–11.
106. Rösche J, Mann K, Fell J (1994): Nonlinear EEG dynamics during sleep in depression and schizophrenia. *Int J Neurosci* 75:271–284.
107. Rösche J, Fell J, Beckmann P (1995): Nonlinear analysis of sleep EEG data in schizophrenia: Calculation of the principal Lyapunov exponent. *Psychiatry Res* 56:257–269.
108. Bob P, Chladek J, Susta M, Glaslova K, Jagla F, Kukleta M (2007): Neural chaos and schizophrenia. *Gen Physiol Biophys* 26:298–305.
109. Bob P, Susta M, Chladek J, Glaslova K, Palus M (2009): Chaos in schizophrenia associations, reality or metaphor? *Int J Psychophysiol* 73:179–185.
110. Bonsall MB, Wallace-Hadrill SM, Geddes JR, Goodwin GM, Holmes EA (2011): Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder. *Proc Biol Sci* 279:916–924.
111. Gottschalk A, Bauer MS, Whybrow PC (1995): Evidence of chaotic mood variation in bipolar disorder. *Arch Gen Psychiatry* 52:947–959.
112. Wichers M, Wigman J, Myin-Germeys I (2015): Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Rev* 7:362–367.
113. Martignoli S, Stoop R (2008): Phase-locking and Arnold coding in prototypical network topologies. *Discrete Continuous Dyn Syst Ser B* 9:145.
114. Naze S, Bernard C, Jirsa V (2015): Computational modeling of seizure dynamics using coupled neuronal networks: Factors shaping epileptiform activity. *PLoS Comput Biol* 11:e1004209.
115. Jirsa VK, Stacey WC, Quilichini PP, Ivanov AI, Bernard C (2014): On the nature of seizure dynamics. *Brain* 137:2210–2230.
116. Izhikevich EM (2007): *Dynamical Systems in Neuroscience*. Cambridge, MA: MIT Press.
117. Brunel N (2000): Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J Comput Neurosci* 8:183–208.
118. Durstewitz D (2009): Implications of synaptic biophysics for recurrent network dynamics and active memory. *Neural Netw* 22:1189–1200.
119. Izhikevich EM (2003): Simple model of spiking neurons. *IEEE Trans Neural Netw* 14:1569–1572.
120. Rinzel J, Ermentrout GB (1998): Analysis of neural excitability and oscillations. In: Koch C, Segev I, editors. *Methods in Neuronal Modeling*, 2nd ed. Cambridge, MA: MIT Press, 251–292.
121. Durstewitz D, Vittoz NM, Floresco SB, Seamans JK (2010): Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* 66:438–448.
122. Spiegel DR, Smith J, Wade RR, Cherukuru N, Ursani A, Dobruskina Y, *et al.* (2017): Transient global amnesia: Current perspectives. *Neuropsychiatr Dis Treat* 13:2691.
123. Normann HK, Asplund K, Karlsson S, Sandman PO, Norberg A (2006): People with severe dementia exhibit episodes of lucidity. A population-based study. *Journal of Clinical Nursing* 15:1413–1417.
124. Ramirez-Mahaluf JP, Roxin A, Mayberg HS, Compte A (2015): A computational model of major depression: The role of glutamate dysfunction on cingulo-frontal network dynamics. *Cereb Cortex* 27:660–679.
125. Mayberg HS, Lozano AM, Voon V, McNeely HE, Seminowicz D, Hamani C, *et al.* (2005): Deep brain stimulation for treatment-resistant depression. *Neuron* 45:651–660.
126. UK ECT Review Group (2003): Efficacy and safety of electroconvulsive therapy in depressive disorders: A systematic review and meta-analysis. *Lancet* 361:799–808.

127. van de Leemput IA, Wichers M, Cramer AO, Borsboom D, Tuerlinckx F, Kuppens P, *et al.* (2014): Critical slowing down as early warning for the onset and termination of depression. *Proc Natl Acad Sci U S A* 111:87–92.
128. Sakoğlu Ü, Pearlson GD, Kiehl KA, Wang YM, Michael AM, Calhoun VD (2010): A method for evaluating dynamic functional network connectivity and task-modulation: Application to schizophrenia. *Magn Reson Mat Phys Biol Med* 23:351–366.
129. Kaiser RH, Whitfield-Gabrieli S, Dillon DG, Goer F, Beltzer M, Minkel J, *et al.* (2016): Dynamic resting-state functional connectivity in major depression. *Neuropsychopharmacology* 41:1822–1830.
130. Jones DT, Vemuri P, Murphy MC, Gunter JL, Senjem ML, Machulda MM, *et al.* (2012): Non-stationarity in the “resting brain’s” modular architecture. *PLoS One* 7:e39731.
131. Uhlhaas PJ, Singer W (2010): Abnormal neural oscillations and synchrony in schizophrenia. *Nat Rev Neurosci* 11:100.
132. Flor-Henry P, Yeudall L, Koles Z, Howarth B (1979): Neuropsychological and power spectral EEG investigations of the obsessive-compulsive syndrome. *Biol Psychiatry* 14:119–130.
133. Grin-Yatsenko VA, Baas I, Ponomarev VA, Kropotov JD (2009): EEG power spectra at early stages of depressive disorders. *J Clin Neurophysiol* 26:401–406.
134. Brodersen KH, Deserno L, Schlagenhaut F, Lin Z, Penny WD, Buhmann JM, *et al.* (2014): Dissecting psychiatric spectrum disorders by generative embedding. *Neuroimage Clin* 4:98–111.
135. Schlösser RG, Wagner G, Koch K, Dahnke R, Reichenbach JR, Sauer H (2008): Fronto-cingulate effective connectivity in major depression: A study with fMRI and dynamic causal modeling. *Neuroimage* 43:645–655.
136. Schlösser RG, Wagner G, Schachtzabel C, Peikert G, Koch K, Reichenbach JR, *et al.* (2010): Fronto-cingulate effective connectivity in obsessive compulsive disorder: A study with fMRI and dynamic causal modeling. *Hum Brain Mapp* 31:1834–1850.
137. Deserno L, Sterzer P, Wüstenberg T, Heinz A, Schlagenhaut F (2012): Reduced prefrontal-parietal effective connectivity and working memory deficits in schizophrenia. *J Neurosci* 32:12–20.
138. Schlagenhaut F, Huys QJM, Deserno L, Rapp MA, Beck A, Heinz H-J, *et al.* (2014): Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage* 89:171–180.
139. Demanuele C, Bähler F, Plichta MM, Kirsch P, Tost H, Meyer-Lindenberg A, *et al.* (2015): A statistical approach for segregating cognitive task stages from multivariate fMRI BOLD time series. *Front Hum Neurosci* 9:537.
140. Koppe G, Toutounji H, Kirsch P, Lis S, Durstewitz D (2019): Identifying nonlinear dynamical systems via generative recurrent neural networks with applications to fMRI. *PLoS Comput Biol* 15:e1007263.
141. Jensen HJ (1998): *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems*. Cambridge, UK: Cambridge University Press.
142. Nonnenmacher M, Behrens C, Berens P, Bethge M, Macke JH (2017): Signatures of criticality arise from random subsampling in simple population models. *PLoS Comput Biol* 13:e1005718.
143. Aksay E, Gamkrelidze G, Seung H, Baker R, Tank D (2001): In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nat Neurosci* 4:184–193.
144. Inagaki HK, Fontolan L, Romani S, Svoboda K (2019): Discrete attractor dynamics underlies persistent activity in the frontal cortex. *Nature* 566:212–217.
145. Toutounji H, Durstewitz D (2018): Detecting multiple change points using adaptive regression splines with application to neural recordings. *Front Neuroinform* 12:67.
146. Takens F (1981): Detecting strange attractors in turbulence. In: Rand DA, Young L-S, editors. *Dynamical Systems and Turbulence, Lecture notes in Mathematics*. Berlin, Germany: Springer-Verlag, 366–381.
147. Niessing J, Friedrich RW (2010): Olfactory pattern classification by discrete neuronal network states. *Nature* 465:47–52.
148. Koppe G, Guloksuz S, Reininghaus U, Durstewitz D (2018): Recurrent neural networks in mobile sampling and intervention. *Schizophr Bull* 45:272–276.
149. Durstewitz D (2017): A state space approach for piecewise-linear recurrent neural networks for identifying computational dynamics from neural measurements. *PLoS Comput Biol* 13:e1005542.
150. Brunton SL, Proctor JL, Kutz JN (2016): Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc Natl Acad Sci U S A* 113:3932–3937.
151. Trischler AP, D’Eleuterio GM (2016): Synthesis of recurrent neural networks for dynamical system simulation. *Neural Netw* 80:67–78.
152. Funahashi K-I, Nakamura Y (1993): Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Netw* 6:801–806.
153. Kimura M, Nakano R (1998): Learning dynamical systems by recurrent neural networks from orbits. *Neural Netw* 11:1589–1599.
154. Durstewitz D, Koppe G, Meyer-Lindenberg A (2019): Deep neural networks in psychiatry. *Mol Psychiatry* 24:1583–1598.
155. Kantz H, Schreiber T (2004): *Nonlinear time series analysis*: Cambridge University Press; 2004.
156. Sauer T, Yorke JA, Casdagli M (1991): Embedology. *J Stat Phys* 65:579–616.
157. Wilson HR (1999): *Spikes, Decisions, and Actions: The Dynamical Foundations of Neurosciences*. Oxford, UK: Oxford University Press.
158. Duncker L, Bohner G, Boussard J, Sahani M (2019): Learning interpretable continuous-time models of latent stochastic dynamical systems [published online ahead of print Feb 12]. *arXiv*.
159. Byron MY, Afshar A, Santhanam G, Ryu SI, Shenoy KV, Sahani M (2005): Extracting dynamical structure embedded in neural activity. *Adv Neural Inf Process Syst* 15:1545–1552.
160. Pandarinath C, O’Shea DJ, Collins J, Jozefowicz R, Stavisky SD, Kao JC, *et al.* (2018): Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat Methods* 15:805–815.
161. Roweis S, Ghahramani Z (2001): Learning nonlinear dynamical systems using the expectation-maximization algorithm. In: Haykin S, editor. *Kalman Filtering and Neural Networks*. Hoboken, NJ: John Wiley and Sons, 175–220.
162. Durstewitz D, Seamans JK (2002): The computational role of dopamine D1 receptors in working memory. *Neural Netw* 15:561–572.